

음성신호의 AM-FM 성분 분리를 위한 가변대역폭 필터 구현

이희영*, 김용태

*서울산업대학교 제어계측공학과

TEL: 02-970-6545, FAX: 02-949-2654

한경대학교 제어계측공학과

Realization of Variable Bandwidth Filter for Decomposition of Speech Signals into AM-FM Components

Heyoung Lee*, Youngtae Kim

*Department of Control and Instrumentation,

Seoul National University of Tecnology

E-mail: leehy@snut.ac.kr

Abstract

In this paper, a variable bandwidth filter(VBF) is realized with the purpose of the decomposition of speech signals with time-varying instantaneous frequencies. The proposed VBF can extract AM-FM components of a speech signal whose time-frequency representations(TFRs) are not overlapped in time-frequency domain.

I. 서 론

음성신호에서 흔히 시변 포먼트 주파수(formant frequency) 또는 시변 순시 주파수(instantaneous frequency)를 갖는 AM-FM 신호 성분들을 관찰할 수 있다. 음성신호를 각 AM-FM 성분별로 분리하거나 또는 각 성분별 스펙트럼을 추정할 경우 음성합성, 음성인식 및 음성 데이터 코딩 등에 유용하게 사용할 수 있다 [1, 7, 8, 9]. AM-FM 신호 $s(t) = a(t)\cos(\phi(t))$ 에서 순시 주파수 $\omega(t)$ 는 다음과 같이 정의된다[2, 10, 11].

$$\omega(t) = D\phi(t) \quad (\text{rad/sec}) \quad (1)$$

여기서 $D = d/dt$ 이다. 음성신호를 시간-주파수 영역에서 관찰해 보면 음성신호의 각 성분별로 시변 순시 주파수의 궤적을 따라 에너지가 집중되는 것을 알 수 있다[4].

일반적으로 신호를 시간-주파수 영역에서 관찰 경우 신호에 대한 순시 주파수, 순시 대역폭(instantaneous bandwidth) 및 각 성분별 경계에 대한 정보 등을 얻을 수 있다. Short-time Fourier 변환(STFT) 등과 같은 시간-주파수 해석 방법을 사용하여 얻어진 시간-주파수 표현(TFR)에서 순시 주파수의 궤적을 따라 에너지가 집중된 부분을 배경과 분리할 경우 분리된 시간-주파수 표현에 대응하는 시간 영역의 실제 신호(real signal)가 항상 존재하는 것을 보장할 수 없다[4, 6, 9, 12, 13]. 변형된 시간-주파수 표현(masked TFR)을 역 STFT을 할

경우 일반적으로 시간 영역에서 허수부분이 나타난다.

본 연구에서는 음성신호에서 시변 순시 주파수를 갖는 각 성분들을 분리하기 위한 IIR(infinite impulse response) 특성을 갖는 선형 시변 가변 대역폭 필터(VBF)의 구현 방법을 제안하였다. 제안된 방법을 사용할 경우 시변 순시 주파수를 갖는 신호들로 이루어진 다중 신호를 각 성분별로 효과적으로 분리해 낼 수 있다.

II. 가변 대역폭 필터

다음과 같이 미분 연산자 G 를 정의하자[3, 5].

$$Gx(t) = g(t)Dx(t) \quad (2)$$

여기서 $g(t)$ 는 임의의 시간 함수로써 $0 < g(t) < \infty$ 이다. 미분 연산자 G 를 사용하여 구성한 선형 시변 미분방정식을 고려하자.

$$a_n G^n y + a_{n-1} G^{n-1} y + \dots + a_0 y = b_m G^m x + \dots + b_0 x \quad (3)$$

여기서 a_i, b_i 는 상수이고, x 는 입력, y 는 출력이다. 연산자 G 는 선형이므로, $Gy = \lambda y$ 의 관계로부터 $Gy^m = \lambda^m y^m$ 임을 알 수 있다. 여기서 m 은 양의 정수이고, λ 는 상수이다. 그러므로 연산자 G 에 대한 식 (3)의 특성 방정식(characteristic equation)은 다음과 같다[3, 5].

$$a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0 = 0 \quad (4)$$

(4)의 특성근(characteristic root)이 서로 다를 경우, 근 $\lambda_i = \sigma_i + j\omega_i$ 에 대응하는 고유함수는 $x = c_i \exp((\sigma_i + j\omega_i)t)$

$\int_0^t 1/g(\tau) d\tau$ 이다. 식 (4)의 모든 특성근의 실수부가 음수일 경우, 식 시스템 (3)은 안정하다[5]. 식 (3)에 표시된 시스템의 주파수 해석을 위하여 확장 푸리에 변환(EFT: extended Fourier transform) 및 역 변환을 다음

과 같이 정의하자.

$$F[x(t), g(t)] = \int_{-\infty}^{\infty} x(t) \frac{1}{g(t)} e^{\int_0^t \frac{-j\omega_g}{g(t)} dt} dt = X(\omega_g) \quad (5)$$

$$F^{-1}[X(\omega_g), g(t)] = \int_{-\infty}^{\infty} X(\omega_g) e^{\int_0^t \frac{j\omega_g}{g(t)} dt} d\omega_g = x(t) \quad (6)$$

여기서 ω_g 는 주어진 함수 g 에 대한 확장 푸리에 변수이다. 임의의 신호 $x(t)$ 와 주어진 함수 $g(t)$ 에 대하여

$$F[G^n x(t), g(t)] = (j\omega_g)^n X(\omega_g) \quad (7)$$

이 성립한다. 여기서 n 은 양의 정수이다[21]. 식 (7)로부터 식 (3)은 확장 푸리에 주파수 영역에서 다음과 같이 표현됨을 알 수 있다.

$$Y(\omega_g) = \left(\sum_{k=0}^m b_k (j\omega_g)^k \right) / \left(\sum_{k=0}^n a_k (j\omega_g)^k \right) X(\omega_g) = H(\omega_g) X(\omega_g) \quad (8)$$

여기서 $H(\omega_g)$ 는 식 (3) 시스템의 전달함수이다. 식 (8)에서 주파수 변수 ω_g 의 물리적 의미는 함수 $g(t)$ 에 따라 달라진다[3, 5]. $H(\omega_g)$ 의 차단 주파수는 계수 a_i, b_i 에 의해 결정된다. 식 (6)으로부터 디락 임펄스 함수(Dirac delta function)의 역 확장 푸리에 변환은 다음과 같다.

$$F^{-1}[\delta(\omega_g - \omega_0), g(t)] = e^{\int_0^t \frac{j\omega_0}{g(t)} dt}$$

그러므로 식 (3)과 식 (8)로부터 차단 주파수가 $\bar{\omega}_g$ 인 저역통과 가변 대역폭 필터의 경우 $\omega_0 < \bar{\omega}_g$ 일 때, 상기 신호가 통과함을 알 수 있다.

가변 대역폭 필터의 시간-주파수 영역에서의 특성과 확장 푸리에 주파수 특성을 알아보기 위해 다음과 같이 시간적으로 변하는 진폭을 갖는 선형 칩(chirp) 신호를 고려하자.

$$s(t) = \exp(2 - 2 \sin t \cos t) \cos(8t + 2t^2) \quad (9)$$

식 (9)의 AM-FM 신호의 위상은 $\phi(t) = 8t + 2t^2$ 이므로, 순시 주파수는 $\omega(t) = 8 + 4t$ rad/s 이다. 이상적인 가변 대역폭 필터의 특성을 효과적으로 관찰하기 위하여 시간에 따라 변하는 순시 주파수와 순시 대역폭을 갖는 단일 성분 신호를 대상으로 시뮬레이션을 수행하였다. 그림 1은 신호 $s(t)$ 의 푸리에 변환을 나타낸다.

그림 2의 (a)는 $1/g(t)$ 를 주어진 신호의 순시 대역폭의 상한으로 택하였을 때, 신호 $s(t)$ 의 EFT이다. 또한 그림 2의 (b)는 $1/g(t)$ 를 주어진 신호의 순시 대역폭의 하한으로 택하였을 때, 신호 $s(t)$ 의 EFT이다. 그림 2의 (a)에서 대부분의 스펙트럼이 주파수 범위 $|\omega_g| < 1$ 내에 있음을 알 수 있다. 또한 그림 2의 (b)에서 대부분의 스펙트럼이 주파수 범위 $|\omega_g| > 1$ 내에 있음을 알 수 있다. 그러므로 그림 2의 (a)로부터 차단 주파수가 $\bar{\omega}_g = \pm 1$ 이고 함수 $1/g(t)$ 를 주어진 신호의 순시 대역폭의 상한으로 선택한 저역통과 가변 대역폭 필터에 신호 $s(t)$ 를 적용하면 신호의 순시 대역폭의 위쪽 경계의 바깥 부분에 있는 잡음을 제거할 수 있음을 알 수 있다. 또한 그림 2의 (b)로부터 차단 주파수가 $\bar{\omega}_g = \pm 1$ 이고 함수 $1/g_h(t)$ 를 주어진 신호의 순시 대역폭의 하한으로 선택한 고역통과 가변 대역폭 필터에 신호 $s(t)$ 를 적용하면 신호의 순

시 대역폭의 아래쪽 경계의 바깥부분에 있는 잡음을 제거할 수 있음을 알 수 있다.

그림 3은 그림 1과 같은 대역통과 가변 대역폭 필터를 사용하여 백색 잡음이 섞인 신호 $s(t)$ 의 잡음을 제거한 신호를 나타낸다. 여기서 저역통과 가변 대역폭 필터는 차단 주파수를 $\bar{\omega}_l = \pm 1$ 로 하고 $1/g_l(t)$ 를 신호 $s(t)$ 의 순시 대역폭의 위쪽 경계로 선택하여 구현하였고, 고역통과 가변 대역폭 필터는 차단 주파수를 $\bar{\omega}_h = \pm 1$ 로 하고 $1/g_h(t)$ 를 신호 $s(t)$ 의 순시 대역폭의 아래쪽 경계로 선택하여 구현하였다. 실험에 의하면 이상적인 가변 대역폭 필터를 이용하여 잡음을 제거한 경우 신호 대 잡음비는 21.4 dB이고, 선형 시불변 필터를 이용하여 잡음을 제거한 경우에는 신호 대 잡음비가 13.8 dB이다. 따라서 선형 시불변 필터보다 가변 대역폭 필터의 잡음 제거 성능이 매우 뛰어나다는 것을 알 수 있다.

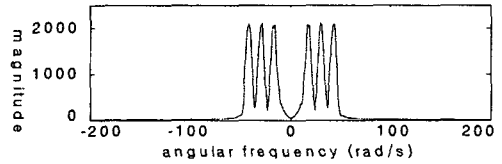


그림 1. 식 (9) 신호의 푸리에 변환

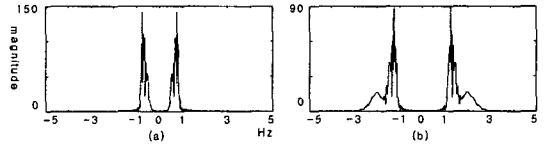


그림 2. 식 (8) 신호의 EFT, x축은 확장 푸리에 변수임. (a) 신호 $s(t)$ 의 순시 대역폭의 상한을 $1/g$ 로 택한 경우. (b) 신호 $s(t)$ 의 순시 대역폭의 하한을 $1/g$ 로 택한 경우

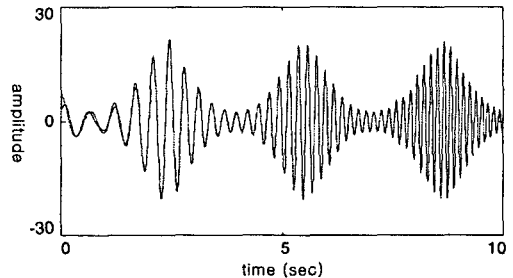


그림 3. 백색 잡음이 있는 신호 $s(t)$ 를 직렬 연결형 대역통과 가변 대역폭 필터를 사용하여 복구한 신호. 실선 : 원 신호, 점선 : 복구한 신호

III. 가변대역폭 필터의 설계 및 구현

식 (3)의 계수 a_i, b_i 에 따라 여러 가지 특성을 갖는 가변대역폭 필터를 구성할 수 있다. 예를 들어, Butterworth 다항식을 사용할 경우 가변대역폭 필터를 구현할 수 있다. 계수의 선택에 따라 저역통과 가변 대

역폭 필터, 고역통과 가변 대역폭 필터, 대역통과 가변 대역폭 필터 및 대역저지 가변 대역폭 필터 등을 설계할 수 있다. 각각의 가변 대역폭 필터들의 차단 주파수는 계수 a_i, b_i 에 의해 결정된다.

가변 대역폭 필터를 시간-주파수 영역에서 신호의 순시 대역폭의 바깥부분에 있는 신호 또는 잡음을 제거하도록 설계할 수 있다. 시간-주파수 영역에서 통과 대역은 차단 주파수 및 $1/g(t)$ 에 의해 결정된다. 저역통과 가변대역폭 필터의 경우 차단 주파수가 $\bar{\omega}_l$ 일 때, 통과 대역과 저지 대역의 경계는 $\bar{\omega}_l/g(t)$ 이다. 시간-주파수 평면 $t-\omega$ 에서 저역통과 가변 대역폭 필터의 통과 영역 R_l 은 다음과 같다.

$$R_l = \{(t, \omega) | \omega < \bar{\omega}_l/g(t), \forall t\} \quad (10)$$

여기서 ω 는 푸리에 주파수 변수이다. 즉, $g(t)=1$ 인 경우의 확장 푸리에 주파수 변수이다. 고역통과 필터의 경우 차단 주파수가 $\bar{\omega}_h$ 일 때, 통과 대역과 저지 대역의 경계는 $\bar{\omega}_h/g_h(t)$ 이다. 시간-주파수 평면 $t-\omega$ 에서 고역통과 가변 대역폭 필터의 통과 영역 R_h 는 다음과 같다.

$$R_h = \{(t, \omega) | \omega > \bar{\omega}_h/g_h(t), \forall t\} \quad (11)$$

함수 $g_b(t)$ 를 가지고 구성된 대역통과 필터에서 차단 주파수가 $\bar{\omega}_a, \bar{\omega}_b$ 일 때 ($\bar{\omega}_a < \bar{\omega}_b$), 통과 영역 R_b 는 다음과 같다.

$$R_b = \{(t, \omega) | \bar{\omega}_a/g_b(t) < \omega < \bar{\omega}_b/g_b(t), \forall t\} \quad (12)$$

식 (10)와 같은 통과 대역을 갖는 저역통과 필터와 식 (11)과 같은 통과 대역을 갖는 고역통과 필터를 직렬 연결하여 구성한 대역통과 필터의 통과 영역 R_{bl} 은 다음과 같다.

$$R_{bl} = \{(t, \omega) | \bar{\omega}_h/g_h(t) < \omega < \bar{\omega}_l/g_l(t), \forall t\} \quad (13)$$

저역통과 및 고역통과 필터를 직렬 연결하여 대역통과 필터를 구성할 경우 통과 대역의 위쪽 경계 및 아래쪽 경계를 임의로 선택할 수 있다. 그러므로 직렬 연결하여 구성된 대역통과 필터의 경우 필터의 잡음 제거 성능이 매우 우수하다.

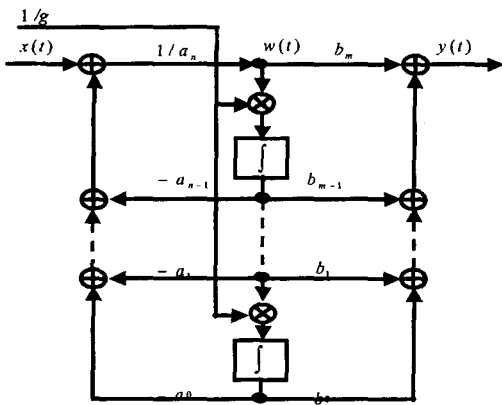


그림 4 직접 표현 II 형태로 구현된 VBR

다음은 가변대역폭 구현 방법을 살펴보자. 연산자 G 와 역 연산자 G^{-1} 는 다음과 같은 관계를 갖는다.

$$GG^{-1}x(t) = G^{-1}Gx(t) = x(t) \quad (14)$$

여기서 G^{-1} 는 다음과 같다.

$$G^{-1}x(t) = \int (1/g(t))x(t)dt \quad (15)$$

그림 4는 식 (14)를 이용하여 식 (3)을 직접 표현 II(direct form II) 형태로 구현한 가변 대역폭 필터다. 식 (15)로부터 G^{-1} 는 곱셈기와 적분기로 구현할 수 있음을 알 수 있다. 곱셈기의 입력 중에 하나는 $1/g(t)$ 이다. $1/g(t)$ 에 의해서 가변 대역폭 필터의 통과 대역 경계가 결정된다. 그림 4는 피드백을 가지고 있으므로 IIR 특성을 갖는다는 것을 알 수 있다.

IV. 음성신호의 AM-FM 성분 분리

그림 5은 성인 남성의 노래의 일부분으로 모음 “아”의 시간 영역 파형이다. 샘플링 주파수가 8 kHz인 16-bit A/D 컨버터를 사용하였다. 그림 6은 그림 5의 시간 $t=0.8$ sec 및 $t=1.3$ sec 사이에 있는 신호의 시간-주파수 표현이다. 여기서 8개의 시변 순시 주파수를 관찰할 수 있다. 각 순시 주파수를 $f_0(t), \dots, f_7(t)$ 라 하자(그림 6참고). 그림 6에 표시된 $b_0(t), \dots, b_6(t)$ 는 각 시변 순시 주파수를 갖는 신호를 분리하기 위해 사용되는 대역통과 가변 대역폭 필터의 통과대역의 상한과 하한이다. 예를 들어, 두 번째 순시 주파수 $f_1(t)$ 를 분리하기 위해 $b_0(t)$ 를 아래쪽 경계로 하고, $b_1(t)$ 을 위쪽 경계로 하는 대역통과 가변 대역폭 필터를 사용한다. 그림 7의 (a)는 그림 5의 시간 $t=0.8$ sec 및 $t=1.3$ sec 사이에 있는 신호의 푸리에 변환이다. 그림 7의 (b)는 $b_3(t)$ 를 $1/g(t)$ 로 택하여 음성신호를 확장 푸리에 변환한 것을 나타낸다. 그림 7의 (a)에서 각 성분별 푸리에 스펙트럼이 서로 겹쳐져 음성신호의 각 성분을 분리하기가 매우 어려움을 알 수 있다. 그러나 그림 7의 (b)에서는 각 순시 주파수 성분이 서로 분리되어 있음을 확인할 수 있다. 그림 8은 대역통과 가변 대역폭 필터를 사용하여 음성신호를 각 성분별로 분리한 신호를 나타낸다.

V. 결론 및 향후 연구

음성신호의 각 성분별 푸리에 스펙트럼이 서로 겹쳐지는 경우에는 기존의 방법을 사용하여 각각의 순시 주파수에 해당되는 신호성분들을 분리하는 것은 매우 어렵다. 본 연구에서는 음성신호에서 시변 순시 주파수를 갖는 각 신호성분을 분리할 수 있는 가변 대역폭 필터 이론 및 구현 방법을 제안하였으며, 이를 사용하여 각 성분별로 음성신호를 분리하였다. 음성신호에서 시변 순시 주파수를 갖는 각 신호 성분을 분리할 경우 음성인식이나 음성코딩에 매우 유용하게 사용할 수 있다. 시간-주파수 영역에서 임의의 통과대역을 갖는 가변 대역폭 필터를 설계할 수 있다. 기존의 방법을 사용할 경우 다중 신호의 각 성분들이 시간-주파수 영역에서 겹쳐지지 않으나 푸리에 주파수 영역에서 겹쳐질 경우 다중 신호를 각 성분별로 분리해 내기가 매우 어렵다. 그러나 이와 같은 신호를 확장 푸리에 주파수 영역에서 관찰할 경우 각 신호성분의 확장 푸리에 스펙트럼은 서로 겹쳐지지 않으므로 각 성분별로 쉽게 분리할 수 있다. 신호의 시간-주파수 표현을 얻기 위해 사용되는 창 함수의 크기에 따라 신호의 스펙트로그램이 달라지므로 가변 대역폭 필터의 성능을 향상시키기 위하여 창 함수의 효과를 제거하기 위한 연구가 필요하다.

참고문헌

- [1] Ashwin Rao, Ramdas Kumaresan, 2000, "On Decomposing Speech into Modulated Components," *IEEE Tr. on Speech and Audio Processing*, Vol. 8, No. 3, 240-254
- [2] Boualem Boashash, 1992, "Time-frequency Signal Analysis," published by Longman Cheshire
- [3] Heyoung Lee, and Zeungnam Bien, 1998, "Reconstruction of Signals with Known Instantaneous Frequency using Linear Time-varying Filter," *Electronics Letters*, Vol. 34, No. 24, 2312-2313
- [4] Shie Qian, and Dapang Chen, 1996, "Joint Time-frequency Analysis: Methods and Applications," published by Prentice-Hall
- [5] Heyoung Lee, and Zeungnam Bien, 1998, "On the Eigenstructure of Linear Quasi-time-invariant Systems," *Int. J. Syst. Sci.*, Vol. 29, No. 8, 873-887
- [6] Griffin, D. W., and Lim, J. S., 1984, "Signal Estimation from Modified Short-time Fourier Transform," *IEEE Tr. on ASSP*, Vol. 32, No. 2, 236-243
- [7] Portnoff, M. R., 1980, "Time-frequency Representation of Digital Signals and Systems based on Short-time Fourier Analysis," *IEEE Tr. on ASSP*, Vol. 28, No. 1, 55-69
- [8] Douglas O'Shaughnessy, 2000, "Speech Communications Human and Machine," published by Piscataway
- [9] T.F. Quatieri, T.E. Hanna, and G.C. O'Leary, 1997, "AM-FM Separation using Auditory Motivated Filters," *IEEE Tr. on Speech Audio Processing*, Vol. 5, No. 4, 465-480
- [10] Boualem Boashash, 1992, "Estimating and Interpreting the Instantaneous Frequency of a Signal-part I, II: Fundamentals," *Proc. IEEE*, Vol. 80, No. 4, 520-568
- [11] Bernard Picinbono, 1997, "On Instantaneous Amplitude and Phase of Signals," *IEEE Tr. on Signal Processing*, Vol. 45, No. 3, 552-560
- [12] L.B. Almeida and J.M. Tribolet, 1983, "Nonstationary Spectral Modeling of Voiced Speech," *IEEE Tr. on ASSP*, Vol. 31, No. 6, 664-678
- [13] R.J. McAulay and T.F. Quatieri, 1986, "Speech Analysis-synthesis based on a Sinusoidal Representation," *IEEE Tr. on ASSP*, Vol. 34, No. 7, 744-754

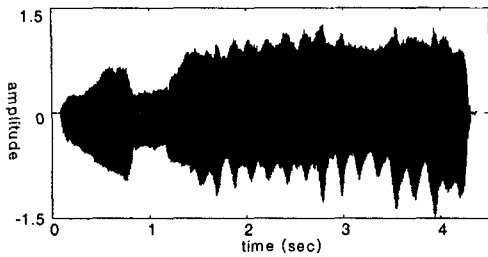


그림 5. 30세 성인 남성의 음성신호(모음 "아")

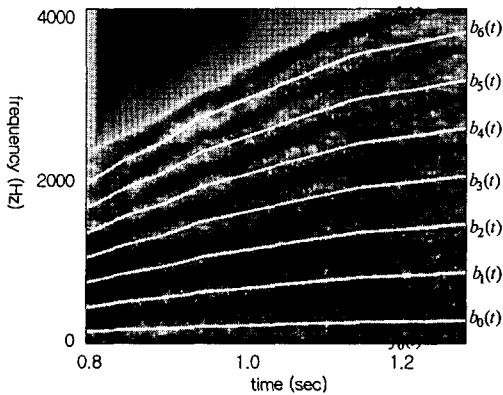


그림 6. 그림 7의 신호의 일부분의 시간-주파수 표현(dB). 백색 실선 : 각 성분별 경계

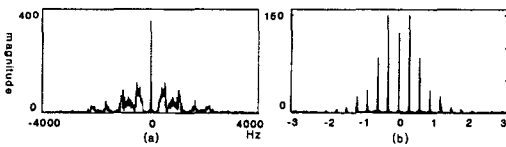


그림 7. 그림 5의 음성신호의 일부분의 주파수 표현(0.8<t<1.3). (a) 푸리에 변환, (b) EFT, $b_3(t)$ 를 $1/g_1(t)$ 로 택한 경우

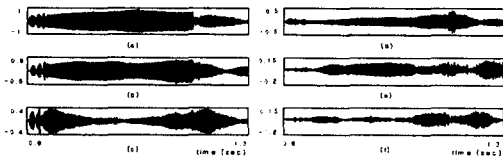


그림 8. 대역통과 가변 대역폭 필터를 사용하여 각 성분별로 분리한 신호. (a) $1/g_1=b_0$, $1/g_1=b_1$, (b) $1/g_1=b_1$, $1/g_1=b_2$, (c) $1/g_1=b_2$, $1/g_1=b_3$, (d) $1/g_1=b_3$, $1/g_1=b_4$, (e) $1/g_1=b_4$, $1/g_1=b_5$, (f) $1/g_1=b_4$, $1/g_1=b_6$